

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

NORMALISASI KATA BAHASA MELAYU MANADO MENGUNAKAN METODE *LEVENSHTEIN DISTANCE*

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik Pada
Jurusan Teknik Informatika



Oleh:

RENGHAD HARTONI
11351102763



**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
PEKANBARU**

2021

LEMBAR PERSETUJUAN

NORMALISASI KATA BAHASA MELAYU MANADO MENGUNAKAN METODE LEVENSHTTEIN DISTANCE

TUGAS AKHIR

Oleh

RENGHAD HARTONI
11351102763

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir

di Pekanbaru, pada tanggal 22 Februari 2021

Pembimbing I,



Muhammad Fikry, S.T., M.Sc
NIP. 19801018 200710 1 002

Pembimbing II,



Yusra, S.T., M.T
NIP. 19840123 201503 2 001

LEMBAR PENGESAHAN

NORMALISASI KATA BAHASA MELAYU MANADO MENGUNAKAN METODE LEVENSHTTEIN DISTANCE

TUGAS AKHIR

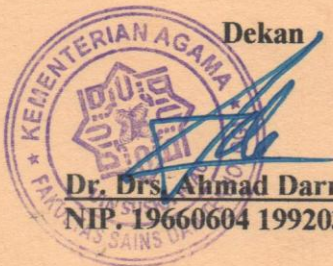
Oleh

RENGHAD HARTONI
11351102763

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
di Pekanbaru, pada tanggal 22 Februari 2021

Pekanbaru, 22 Februari 2021

Mengesahkan,



Dekan

Dr. Drs. Ahmad Darmawi, M.Ag.
NIP. 19660604 199203 1 004

Ketua Jurusan

Dr. Elin Hadiani, S.T., M.Kom.
NIP. 19810523 200710 2 003

DEWAN PENGUJI

Ketua : Dr. Alwis Nazir, M.Kom.

Sekretaris : Muhammad Fikry, S.T., M.Sc.

Anggota I : Yusra, S.T., M.T.

Anggota II : Iwan Iskandar, M.T.

Anggota III: Suwanto Sanjaya, S.T., M.Kom.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan seizin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi nama, tanda peminjaman, dan tanggal pinjam.

UIN SUSKA RIAU



LEMBAR PERNYATAAN

Dengan ini menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar keserjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu naskah ini dan disebutkan di dalam daftar pustaka.

Pekanbaru, 26 Februari 2021

Yang membuat pernyataan,

Renghad Hartoni
11351102763

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

KATA PENGANTAR

Puji Syukur kepada Tuhan YME yang senantiasa memberikan rahmat dan karunia-Nya, sehingga Tugas Akhir yang berjudul “Normalisasi Kata Bahasa Melayu Manado Menggunakan Metode *Levenshtein Distance*” dapat diselesaikan.

Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana strata satu pada Program Studi Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau. Selama penyusunan Tugas Akhir, penulis sangat banyak mendapatkan pengalaman, pengetahuan, bimbingan dan tunjuk ajar, serta dukungan dari berbagai pihak yang telah membantu hingga Tugas Akhir ini dapat diselesaikan. Untuk itu pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Bapak Prof.Dr. Suyitno, M.Ag selaku Rektor UIN SUSKA Riau.
2. Bapak Dr. Drs. Ahmad Darmawi, M.Ag., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Ibu Dr. Elin Haerani, S.T., M.Kom. selaku Ketua Jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau
4. Ibu Fadhila Syafria, S.T, M.Kom selaku Koordinator Tugas Akhir Jurusan Teknik Informatika.
- Ibu Yusra, S.T., M.T., selaku Pembimbing Akademis dan Pembimbing II Tugas Akhir penulis yang memberikan arahan akademis selama menempuh perkuliahan.
- Bapak Muhammad Fikry, S.T., M.Sc., selaku Pembimbing I Tugas Akhir penulis yang memberikan arahan dan wawasan dalam penulisan Tugas Akhir.
- Bapak Iwan Iskandar, M.T., selaku Dosen Penguji I Tugas Akhir penulis, yang telah bersedia dan meluangkan waktunya untuk melakukan koreksi kesalahan dalam penulisan tugas akhir ini.
- Suwanto Sanjaya, S.T., M.Kom., selaku Dosen Penguji II Tugas Akhir penulis, yang telah bersedia dan meluangkan waktunya untuk melakukan koreksi kesalahan dalam penulisan tugas akhir ini.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Bapak dan Ibu dosen Teknik Informatika yang Sudah memberikan ilmu pengetahuan kepada penulis.

6. Orang tua penulis, Bapak Jetro Silalahi dan Ibu Jelita Sihombing yang senantiasa mendoakan, memberikan nasihat serta dukungan dalam segala aspek.

7. Kepada saudara dan saudari penulis Adre Silalahi, Juliana Silalahi dan Ifran Silalahi.

8. Terima kasih juga kepada kawan-kawan seperjuangan TIF B 2013 “PLAN B”.

13. Kepada kawan-kawan satu jurusan yang telah sama-sama berbagi ilmu, informasi serta keluh kesah selama proses pembuatan Tugas Akhir.

Dan semua pihak yang terlibat baik langsung maupun tidak langsung dalam pengerjaan tugas akhir ini yang tidak dapat penulis sebutkan namanya satu persatu.

Semoga penelitian ini dapat bermanfaat bagi penulis khususnya dan untuk pembaca pada umumnya. Serta dapat menjadi referensi dan rujukan bagi hal-hal yang bermanfaat. Sebagai manusia tentunya penulis tidak lepas dari kekurangan dan kesalahan, baik dalam pelaksanaan penelitian maupun dalam penyusunan Laporan Tugas Akhir ini. Oleh karena itu, penulis menyampaikan permohonan maaf yang sebesar-besarnya. Penulis berharap adanya kritik dan saran guna memperbaiki atau sebagai pengembangan kedepannya. Kritik dan saran tersebut dapat dikirim ke email penulis: renghad.hartoni@students.uin-suska.ac.id. Akhir kata penulis ucapkan terima kasih.

Pekanbaru, 10 Januari 2021

UIN SUSKA RIAU

Penulis



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

NORMALISASI KATA BAHASA MELAYU MANADO MENGUNAKAN METODE *LEVENSHTEIN DISTANCE*

RENGHAD HARTONI

11351102763

Tanggal Sidang :

Periode Wisuda :

Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRAK

Bahasa Melayu Manado adalah bahasa yang dipertuturkan oleh masyarakat Sulawesi Utara sebagai bahasa pemersatu bangsa. Bahasa Melayu Manado masih digunakan hingga saat ini termasuk di media sosial. Seperti pada penggunaan bahasa pada umumnya, ada kemungkinan terjadinya kesalahan dalam penulisan ataupun pengetikan. Penelitian ini bertujuan untuk menormalisasikan kata yang terdapat kesalahan penulisan dengan memberikan rekomendasi kata yang memiliki jarak kemiripan terkecil dari kata yang salah. Kata yang menjadi rekomendasi berupa kata dasar dan atau kombinasi kata dasar dengan imbuhan yang ada dalam aturan bahasa Melayu Manado. Kata yang memiliki kesalahan penulisan (kata asal) akan di hitung jarak kemiripannya dengan kata target yaitu semua kata dasar dalam kamus serta kata kombinasi kata dasar dengan imbuhan Bahasa Melayu Manado. Keseluruhan proses diawali dengan melakukan tahap *preprocessing* yaitu *cleaning*, *case folding* dan *tokenizing* pada data uji. Selanjutnya dilakukan proses *stemming* terhadap data uji yang bertujuan untuk menghapuskan imbuhan yang ada pada data uji. Proses normalisasi dilakukan kepada data uji yang tidak memiliki kecocokan pada kamus setelah proses *stemming*. Metode *Levenshtein Distance* berperan untuk menghitung jarak kemiripan antara dua buah data *string* dengan melakukan proses penyisipan, penghapusan dan penggantian karakter. Penghitungan akurasi penelitian ini menghasilkan 35.8%.

Kata Kunci : Bahasa Melayu Manado, *Levenshtein Distance*, Normalisasi Kata, *Stemming*, *Preprocessing*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

MANADO MALAY WORDS NORMALIZATION USING LEVENSHTEIN DISTANCE METHOD

RENGHAD HARTONI

11351102763

Date of Final Exam :

Graduation Ceremony Period :

Informatics Engineering Department

Faculty of Science and Technology

Islamic State University of Sultan Syarif Kasim Riau

ABSTRACT

Manado Malay or Manado Language is the language spoken by the people in North Sulawesi as the unifying language of the nation. Manado Malay is still used until today, including on social media. As with the use of language in general, there is the possibility of errors in writing or typing. This study aims to normalize words that have writing errors by providing word recommendations that have the smallest similarity distance from the incorrect words. The word that becomes the recommendation is in the form of a root word and or a combination of root words with affixes that are exist in the rules of the Manado Malay language. The similarity distance between the word that has errors (origin word) and the target words which are the root words in the dictionary including the combination words with the affixes that are exist in Manado Malay will be calculated. The whole process begins with the preprocessing step which are cleaning, case folding and tokenizing on the data test. Furthermore, the stemming process is carried out on the data test which aims to eliminate the affixes in the data test. The normalization process was carried out on the data test which did not match the dictionary after the stemming process. The Levenshtein Distance method is used to calculate the similarity distance between two data strings by performing the process of character insertion, deletion and replacement. The calculation result of the accuracy of this research is 61.97%.

Keyword : Levenshtein Distance, Manado Malay Preprocessing,, Stemming, Word Normalization



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR ISI

LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN	I-i
LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL	I-i
LEMBAR PERNYATAAN.....	I-i
KATA PENGANTAR.....	I-i
ABSTRAK	I-i
DAFTAR ISI	I-i
DAFTAR GAMBAR.....	I-v
DAFTAR TABEL	I-vi
DAFTAR RUMUS	I-i
DAFTAR SIMBOL.....	1
BAB I PENDAHULUAN	I-1
1.1 Latar Belakang	I-1
1.2 Rumusan Masalah	I-2
1.3 Batasan Masalah.....	I-3
1.4 Tujuan Penelitian.....	I-3
1.5 Sistematika Penulisan.....	I-3
BAB II LANDASAN TEORI	II-1
2.1 Natural Language Processing (NLP).....	II-1
2.2 Text Mining.....	II-1
2.3 <i>Text Pre-processing</i>	II-2

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.4	Normalisasi Kata	II-3
2.5	Levenshtein Distance	II-3
2.6	Bahasa Melayu Manado	II-4
2.7	Imbuhan Bahasa Melayu Manado	II-5
2.8	Perhitungan Akurasi (<i>Suggestion Adequacy</i>)	II-12
2.9	Penelitian Terkait	II-13

BAB III METODOLOGI PENELITIAN III-1

3.1	Perumusan Masalah.....	III-1
3.2	Studi Literatur.....	III-2
3.3	Pengumpulan Data	III-2
3.4	Analisa Algoritma	III-2
3.4.1	Analisa Kebutuhan Data	III-2
3.4.2	Analisa Tahapan.....	III-3
3.5	Perancangan.....	III-3
3.5.1	Perancangan Basis Data	III-4
3.5.2	Perancangan Antarmuka	III-4
3.5.3	Flowchart	III-4
3.6	Implementasi	III-4
3.7	Pengujian	III-5
3.8	Kesimpulan dan Saran.....	III-5

BAB IV ANALISA DAN PERANCANGANIV-1

4.1	Analisa.....	IV-1
4.1.1	Analisa Permasalahan	IV-1
4.1.2	Pengumpulan Data	IV-1
4.1.3	Preprocessing.....	IV-4

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4.1.4	<i>Stemming</i>	IV-5
4.1.5	Normalisasi	IV-7
4.2	Perancangan.....	IV-20
4.2.1	<i>Flowchart</i>	IV-20
4.2.2	Perancangan Basis Data	IV-21
4.2.3	Perancangan Struktur Menu	IV-22
4.2.4	Perancangan Antar Muka	IV-22
BAB V	Implementasi dan Pengujian	V-1
5.1	Implementasi	V-1
5.1.1	Lingkungan Implementasi.....	V-1
5.1.2	Batasan Implementasi	V-1
5.2	Implementasi Antarmuka	V-2
5.2.1	Implementasi Antar Muka Halaman Beranda.....	V-2
5.2.2	Implementasi Antar Muka Halaman Data Kamus	V-2
5.2.3	Implementasi Antar Muka Halaman Data Uji	V-3
5.2.4	Implementasi Antar Muka Halaman Uji Normalisasi.....	V-4
5.2.5	Implementasi Antar Muka Halaman Uji Coba.....	V-4
5.3	Pengujian	V-5
BAB VI	Kesimpulan dan Saran.....	VI-1
6.1	Kesimpulan.....	VI-1
6.2	Saran	VI-1
DAFTAR PUSTAKA.....		ix
Lampiran A Data Validator.....		ix
Lampiran B Pengujian Data Uji.....		ix



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR GAMBAR

	Halaman
3.1 Flowchart Metodologi Penelitian.....	III-1
4.1 Profil Akun @TheBAKUSEDU	IV-2
4.2 Ambil Data Uji.....	IV-3
4.3 Data Uji dalam Basis Data.....	IV-3
4.4 Flowchart Analisa Sistem	IV-21
4.5 Struktur Menu	IV-22
4.6 Rancangan Antar Muka Beranda	IV-23
4.7 Rancangan Antar Muka Data Kamus	IV-23
4.8 Rancangan Antar Muka Data Uji.....	IV-24
4.9 Rancangan Antar Muka Uji Normalisasi	IV-24
4.10 Rancangan Antar Muka <i>Stemming</i>	IV-25
4.11 Rancangan Antar Muka Normalisasi	IV-25
4.12 Rancangan Antar Muka Rekomendasi	IV-26
4.13 Rancangan Antar Muka Hitung Matrik	IV-26
5.1 Tampilan Halaman Beranda	V-2
5.2 Tampilan Halaman Data Kamus.....	V-3
5.3 Tampilan Antar Muka Uji Normalisasi	V-3
5.4 Implementasi Antar Muka Uji Normalisasi	V-4
5.5 Implementasi Antar Muka Uji Coba.....	V-5

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR TABEL

Tabel 4.1 Proses <i>Cleaning</i>	IV-4
Tabel 4.2 Proses <i>Case Folding</i>	IV-4
Tabel 4.3 Proses <i>Tokenizing</i>	IV-5
Tabel 4.4 Proses <i>Stemming</i>	IV-6
Tabel 4.5 Indeks Matriks kata "orng" dan "orang"	IV-9
Tabel 4.6 Matriks Jarak "orng" dan "orang"	IV-14
Tabel 4.7 Indeks Matriks kata "orng" dan "dorang"	IV-15
Tabel 4.8 Matriks Jarak kata "orng" dan "dorang"	IV-18
Tabel 4.9 Indeks Matriks kata "dpa" ke "bakudapa"	IV-19
Tabel 4.10 Matrik Jarak kata "dpa" ke "bakuopa"	IV-19
Tabel 4.11 Matrik Jarak kata "dpa" ke "bakudepa"	IV-19
Tabel 4.12 Matrik Jarak kata "dpa" ke "bakupa"	IV-19
Tabel 4.13 Rekomendasi Kata	IV-20
Tabel 4.14 Data Uji	IV-21
Tabel 4.15 Kata Dasar	IV-22
Tabel 5.1 Pengujian Rekomendasi Kata	V-5

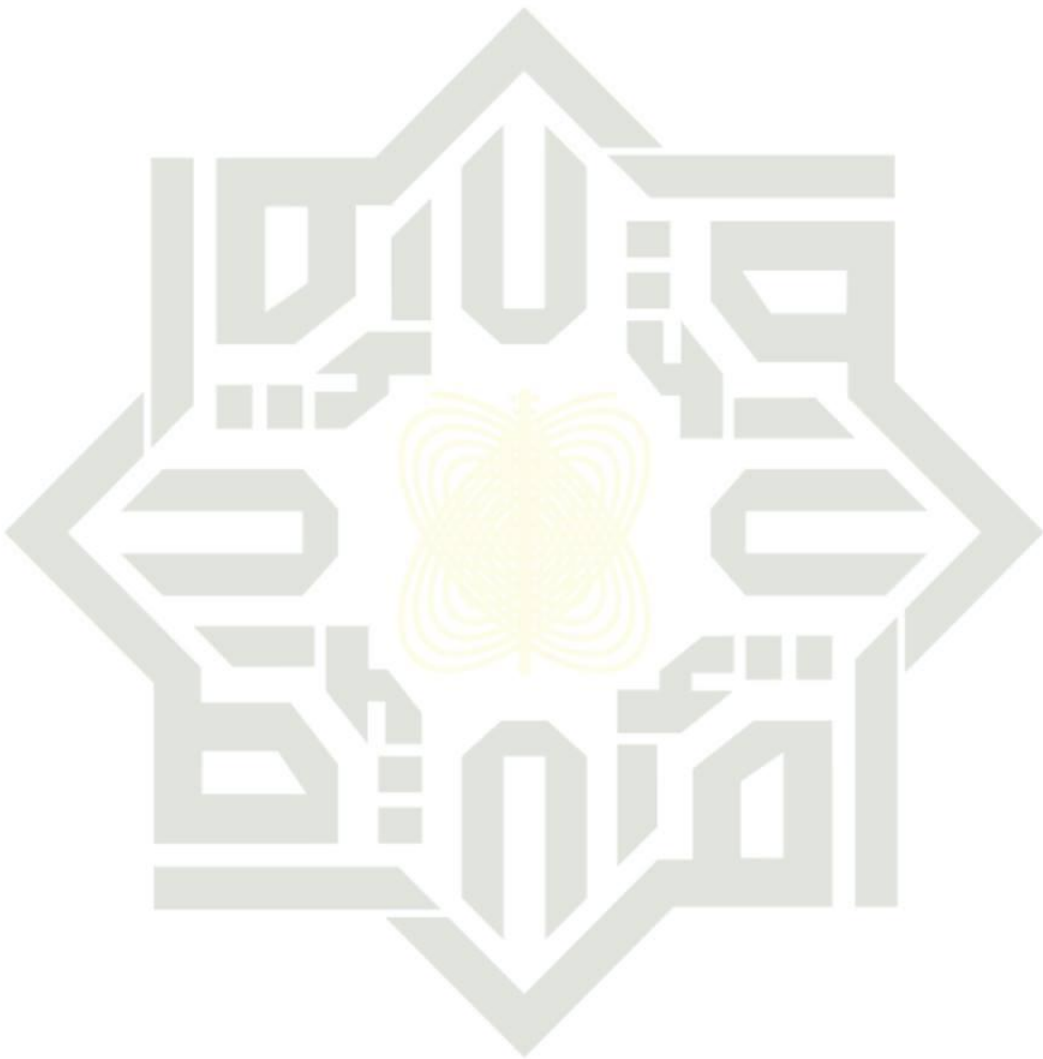


DAFTAR RUMUS

Rumus	Halaman
Rumus Levenshtein Distance	IV-8

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.





Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR SIMBOL

Gambar	Keterangan
	Terminator: Simbol terminator (mulai/ selesai) merupakan tanda bahwa sistem akan dijalankan atau berakhir.
	Proses: Simbol yang digunakan untuk melakukan pemrosesan data baik oleh user maupun komputer (sistem).
	Verifikasi: Simbol yang digunakan untuk memutuskan apakah valid atau tidak validnya suatu kejadian.
	Data: Simbol yang digunakan untuk mendeskripsikan data input/ output yang digunakan.
	Arus Data: Simbol yang digunakan untuk menggambarkan arus data di dalam sistem.
	Display atau aktifitas menampilkan visual oleh layar.
	Prepare: Menyediakan storage penyimpanan dikarenakan adanya lebih dari 1 dokumen yang akan diolah.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB I PENDAHULUAN

1.1 Latar Belakang

Bahasa daerah merupakan kekayaan suatu bangsa karena memiliki nilai-nilai kebudayaan yang terkandung di dalamnya. Kekayaan kebudayaan yang dimiliki Indonesia berupa ragam suku dan budaya. Tercatat sedikitnya Indonesia memiliki 748 bahasa daerah yang menjadi simbol kekayaannya (Widianto, 2018). Beberapa bahasa daerah yang menjadi kekayaan bangsa Indonesia seperti bahasa Lampung, Bahasa Jawa, Bahasa Sunda dan Bahasa Batak. Pada pembahasan ini bahasa daerah yang dijadikan bahan penelitian merupakan bahasa daerah Melayu Manado.

Bahasa Melayu Manado merupakan bahasa yang dipertuturkan oleh masyarakat Manado yang berasal dari Sulawesi Utara. Bahasa Melayu Manado ini masih digunakan sebagai media komunikasi bagi masyarakat Sulawesi Utara. Selain sebagai salah satu pemersatu masyarakat, bahasa Melayu Manado juga merupakan sebuah kebanggaan bagi masyarakat Sulawesi Utara (Najoan, Liwoso, Djojuroto, & Kembuan, 1981).

Sama halnya dengan bahasa yang lain, dalam penulisan bahasa Melayu Manado juga berkemungkinan untuk terjadinya salah ketik (*typo*). *Typo* adalah penulisan kata yang tidak sesuai dengan ejaan kata baku. Penulisan kata yang *typo* mengakibatkan kata memiliki makna yang tidak tepat. Penggunaan kata yang *typo* pada proses klasifikasi (*text mining*) akan membuat analisis menjadi tidak akurat, sehingga diperlukan normalisasi (Zulkifli & Suadaa, 2019).

Normalisasi teks merupakan teknik konversi teks yang tidak baku menjadi kata baku yang sesuai dengan kamus (Rahate & Chandak, 2019). Normalisasi teks kerap kali dimanfaatkan bagi kalangan akademis dan peneliti di bidang *Natural Language Processing* (NLP) yang meneliti informasi-informasi teks di media sosial yang memiliki ejaan tidak baku sehingga menyebabkan disambiguitas makna. Normalisasi menjadikan kinerja NLP efektif dengan cara merepresentasikan kata yang tidak baku menjadi kata baku di dalam satu



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

korpus (Priansya, 2017). Ada beberapa metode yang dapat digunakan dalam proses normalisasi, diantaranya adalah metode *Levenshtein Distance*.

Levenshtein Distance adalah algoritma yang diciptakan oleh Vladimir Levenshtein dimana algoritma ini dapat menentukan jarak (*distance*) dari kemiripan antara dua buah *string*. Jarak kemiripan antar *string* dapat dihitung berdasarkan perubahan antar dua buah kumpulan karakter seperti penambahan, pengurangan atau penggantian karakter (Antinasari, 2017).

Menurut (Nugraha & Rizqullah, 2019) pada penelitian “Normalisasi Kata Tidak Baku yang Tidak Disingkat dengan Jarak Perubahan” menunjukkan bahwa normalisasi dengan jarak *Levenshtein* mengungguli normalisasi dengan jarak *Longest Common Subsequence* (LCS) dengan selisih akurasi sebesar 8,34%. Penelitian terkait lainnya yang menggunakan algoritma *Levenshtein Distance* yaitu penelitian (Pratama & Pamungkas, 2016) tentang “Analisis Kinerja Algoritma *Levenshtein Distance* Dalam Mendeteksi Kemiripan Dokumen Teks” diperoleh hasil simulasi menunjukkan bahwa penggunaan *sorting* sangat berpengaruh bagi algoritma *Levenshtein Distance* yang menggunakan *stopword removal*, *stemming* dan *sorting* sekaligus.

Berdasarkan permasalahan di atas solusi yang dapat dilakukan adalah melakukan normalisasi kata Bahasa Melayu Manado menggunakan *Levenshtein Distance* yang bertujuan untuk memulihkan kata yang mengalami kesalahan penulisan menjadi kata baku dengan memberikan rekomendasi kata dari kamus yang memiliki jarak edit terkecil. Pentingnya dilakukan penelitian tentang normalisasi pada bahasa Melayu Manado agar dapat dilanjutkan dan dikembangkan ketahap selanjutnya yaitu klasifikasi.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, rumusan masalah yang dapat diambil yaitu “Bagaimana menerapkan algoritma *Levenshtein Distance* untuk normalisasi kata pada bahasa Melayu Manado”.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1.3 Batasan Masalah

Supaya pembahasan penelitian ini tidak menyimpang dari apa yang telah dirumuskan, maka diperlukan batasan. Batasan dalam penelitian ini dapat dijabarkan sebagai berikut:

1. Data dalam penelitian ini berupa kamus bahasa Melayu Manado, serta kumpulan *tweet* dari akun *twitter* berbahasa Melayu Manado dengan nama @TheBAKUSEDU.
2. Data acuan (kamus) bersumber dari penelitian sebelumnya di tahun 2019 yaitu algoritma *stemming* teks bahasa Melayu Manado berbasis aturan tata bahasa oleh Rahmad Harahap.
3. Data uji yang akan digunakan merupakan kalimat berbahasa Melayu Manado yang terdapat kesalahan dalam penulisan (*typo*). Data keluaran yang akan dihasilkan berupa kata-kata dengan jarak *levenshtein distance* terkecil.
4. Penggunaan kata gaul dalam *tweet* akan diabaikan dan tidak diproses.

1.4 Tujuan Penelitian

Tujuan dari penelitian tugas akhir ini adalah untuk menerapkan dan menguji algoritma *Levenshtein Distance* untuk normalisasi kata Bahasa Melayu Manado yang mengalami kesalahan penulisan.

1.5 Sistematika Penulisan

Sistematika penulisan penelitian Tugas Akhir yang akan dibuat adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan latar belakang permasalahan, rumusan masalah, batasan masalah, tujuan penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini menjelaskan dasar teori-teori yang digunakan sebagai landasan dalam pembuatan Tugas Akhir. Pengertian normalisasi, sosial media, *stemming*, pengertian *Natural Language Processing*, algoritma *Levenshtein Distance* dan penelitian terkait.



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metodologi yang digunakan dalam penelitian yaitu: Rumusan Masalah, Studi Literatur, Pengumpulan Data, Analisa dan Perancangan, Implementasi dan Pengujian serta Kesimpulan dan Saran.

BAB IV ANALISA DAN PERANCANGAN

Bab ini membahas analisis langkah kerja dan perancangan Algoritma yang akan dibuat sesuai dengan metodologi yang digunakan.

BAB V IMPLEMENTASI DAN PENGUJIAN

Bab ini membahas tentang implementasi dan pengujian hasil dari penelitian sesuai dari rumusan-rumusan dan pembahasan.

BAB VI PENUTUP

Penutup merupakan jawaban atas rumusan masalah dalam penelitian, yang membahas tentang kesimpulan dan saran untuk penelitian selanjutnya.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB II

LANDASAN TEORI

2.1 Natural Language Processing (NLP)

Natural Language Processing (NLP) atau yang dalam bahasa Indonesia disebut Bahasa Pemrosesan Alami adalah serangkaian teknik untuk analisis otomatis dan representasi bahasa manusia (Cambria & White, 2014). NLP yang merupakan salah satu cabang kecerdasan buatan yang berhubungan dengan interaksi antar manusia dan komputer. Dengan kata lain NLP bertujuan untuk membuat mesin memiliki pemahaman bahasa manusia dan kemudian dapat memberikan tanggapan yang sesuai (Alamanda, Suhery, & Brianorman, 2016).

2.2 Text Mining

Text mining merupakan proses menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan Analisa keterhubungan antar dokumen (Putri, 2014).

Secara umum, *text mining* harus melewati beberapa tahapan proses untuk memperoleh hasil yang maksimal. Adapun proses-proses yang biasa dilakukan dalam tahapan ini adalah sebagai berikut:

1. Text Pre-processing

Proses pengubahan bentuk data yang belum terstruktur menjadi data yang terstruktur sesuai dengan kebutuhan untuk proses *text mining* yang lebih lanjut.

2. Text Transformation

Proses transformasi teks melakukan perubahan kata menjadi kata dasar (*stemming*), pengurangan dimensi kata pada dokumen (*stop word removal*), *bag of word*, dan *vectorial document representation* untuk representasi dokumen.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3. *Feature Selection*

Tahapan ini dilakukan pembuangan bagian-bagian yang tidak terkait, *feature* yang tidak berpengaruh akan dibuang dan memfokuskan *feature* yang berpengaruh dalam suatu kegiatan pemodelan atau penganalisaan data.

4. *Pattern Discovery*

Pattern Discovery bertujuan untuk menemukan pola yang tepat sebagai pengetahuan dari teks dengan menggunakan Teknik-teknik *data mining* seperti, *classification* dan *clustering*.

5. *Interpretation*

Tahapan ini melakukan interpretasi ke suatu bentuk untuk kemudian dilakukan evaluasi.

2.3 *Text Pre-processing*

Pre-processing data merupakan proses untuk mempersiapkan data mentah sebelum dilakukan proses lain. Pada umumnya, *text pre-processing* data dilakukan dengan cara mengeliminasi data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah diproses oleh sistem (Mujilahwati, 2016).

Terdapat beberapa tahapan dalam *text pre-processing*, yaitu :

Case Folding, yaitu tahapan mengubah karakter dalam kata yang memiliki huruf kapital menjadi huruf kecil. Dalam tahapan ini karakter yang dapat diubah adalah abjad a-z, sedangkan karakter lain tidak dianggap (Darujati & Bimo Gumelar, 2016).

Cleaning, merupakan tahapan membersihkan teks dengan menghilangkan karakter yang bukan merupakan abjad. Karakter yang dimaksud berupa koma (,), titik (.), strip (-), tanda seru (!) dan lain-lain.

Tokenizing, yaitu tahapan merubah kalimat menjadi kata. Teks yang didapatkan dari media sosial akan dikonversi menjadi string.

Stemming, yaitu proses pengubahan variasi kata, baik kata berimbuhan maupun kata yang berulang menjadi kata dasarnya. Stemming juga

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

berfungsi untuk mencegah ketidakcocokan kata yang tersedia dalam kamus dengan kata yang diinginkan (Sharma, 2012).

2.4 Normalisasi Kata

Normalisasi kata adalah proses mengubah kata yang tidak baku menjadi kata baku dalam suatu bahasa. Normalisasi kata berfungsi untuk menghindari ketiadaan kata dalam kamus. Ketiadaan kata dalam kamus dapat dikarenakan penulisan kata yang disingkat maupun penulisan kata yang salah. Normalisasi meliputi pengubahan singkatan, akronim, angka, tanggal, waktu, karakter-karakter khusus, dan simbol-simbol dengan bentuk huruf alphabet lengkap sehingga tidak terjadi ambiguitas berkenaan dengan cara pengucapan (Purwadi, 2011).

2.5 Levenshtein Distance

Levenshtein Distance adalah metode programming dinamis yang mengatasi permasalahan pola kecocokan atau kemiripan berdasarkan operasi sederhana seperti penyisipan karakter, pemotongan karakter dan pergantian karakter dengan karakter lain (Lyras, Sgarbas, & Fakotakis, 2007).

Terdapat 3 (tiga) operasi dalam algoritma *Levenshtein Distance*, yaitu:

1. Penyisipan

Operasi penyisipan memiliki arti menambahkan karakter kedalam sebuah string. Penambahan karakter dapat terjadi di awal string, di tengah string maupun di akhir string.

2. Penghapusan

Operasi penghapusan memiliki arti menghilangkan karakter yang berada di dalam string.

3. Penggantian

Operasi penggantian memiliki arti mengubah karakter yang menjadi penyusun sebuah string dengan karakter lain yang berbeda.

Operasi-operasi yang dilakukan tersebut dapat dilihat pada Persamaan (2.1).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$D_{(s,t)} = \begin{cases} \min D(s-1, t) + 1 \text{ (Penghapusan)} \\ \min D(s, t-1) + 1 \text{ (Penyisipan)} \\ \min D(s-1, t-1) + 1(a_i \neq b_j) \text{ (Penukaran)} \\ \min D(s-1, t-1) s_j = t_i \text{ (Tidak ada perubahan)} \end{cases}$$

Persamaan 2. 1 Rumus Levenshtein Distance

Keterangan :

s = string sumber

t = string target

D = jarak Levenshtein Distance

s(j) = karakter string sumber ke-j

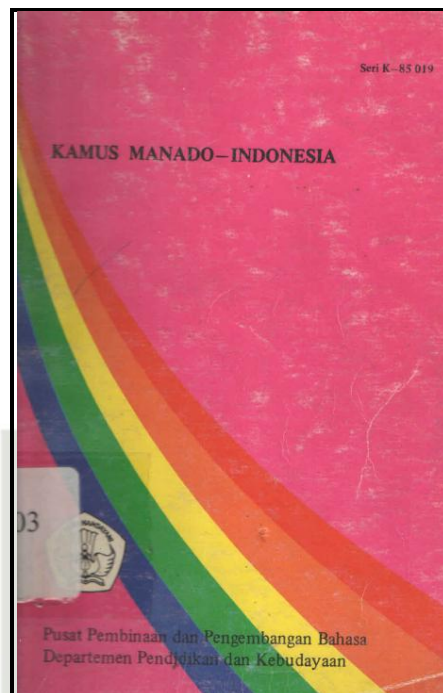
t(i) = karakter string target ke-i

2.6 Bahasa Melayu Manado

Bahasa Melayu Manado digunakan untuk berkomunikasi dalam hampir semua aktivitas kehidupan di Sulawesi Utara. Bahasa Melayu Manado sebagai bahasa perhubungan antaretnis di Sulawesi Utara. Kedudukan bahasa Melayu Manado dikategorikan sebagai bahasa regional di Sulawesi Utara sama seperti bahasa Melayu Betawi di Jakarta. Pengaruh bahasa Melayu Manado sangat kuat sehingga diperkirakan akan menyebabkan bahasa daerah di Sulawesi Utara terancam. Hal ini memunculkan masalah tersendiri mengenai bahasa daerah, khususnya di Minahasa dan umumnya di Sulawesi Utara. Pengelola pendidikan atau pengajaran di Sulawesi Utara harus memperhatikan hal ini. Hal ini disebabkan oleh bahasa Melayu Manado kadang-kadang dalam situasi resmi (dunia pendidikan) dan apalagi tidak resmi di luar sekolah senantiasa tampak pengaruh bahasa ini. Oleh karena itu, bahasa Melayu Manado dapat menggeser bahasa daerah yang ada dan memengaruhi bahasa Indonesia khususnya dalam dunia pengajaran di Sulawesi Utara (Asri, 2014).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 2.II1 Kamus Kata Dasar

2.7 Imbuhan Bahasa Melayu Manado

Imbuhan termasuk dalam tataran bidang ilmu morfologi, dimana morfologi merupakan cabang ilmu bahasa yang mempelajari seluk beluk kata, perubahan kata dan dampak dari perubahan itu terhadap arti dan kelas kata.

Pembentukan kata dilakukan dengan penambahan afiks di depan kata, dibelakang, dan ditengah kata dasar atau akar-akar kata yang masing-masing disebut dengan prefiks, suffiks, dan infiks. Selain itu, pembentukan kata dapat pula dilakukan dengan menambahkan afiks yang berbentuk unsur suprasegmental atau modifikasi yang dilakukan pada salah satu unsur dasar, atau dengan pembaruan unsur segmental pada kata dasar atau akar kata (Najoan et al., 1981). Aturan pembentukan imbuhan dapat dilihat pada Tabel 2.1 berikut.

Tabel 2. 1 Aturan Pembentukan Imbuhan

No	Prefiks	Sufiks	Reduplikasi	Kombinasi Afiks	Konfiks
1	Ba-	-an	Reduplikasi Kata	Ba+baku	Ka- +-an
2	Bar-	-akang	-	Ta+Baku	Pan-+-an
3	Baku-	-	-	Ta+ka	Pa-+ -an
4	Ta-	-	-	Pam+ba	Pam-+ -an

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

No	Prefiks	Sufiks	Reduplikasi	Kombinasi Afiks	Konfiks
5	Ka-	-	-	Pang+ba	Pang- + -an
6	Pan-	-	-	-	-
7	Pa-	-	-	-	-
8	Pam-	-	-	-	-
9	Pang-	-	-	-	-
10	Tar-	-	-	-	-

Berikut merupakan penjabaran Tabel 2.1 di atas :

Prefiks (Awalan)

Imbuhan awalan (Prefiks) digunakan dalam bahasa Melayu Manado terdiri dari (ba-,bar-,baku-,ta-,ka- dan pan-) (Ny.J.A.Karisoh Najoan, M.A Liwoso, Kinajati Djojuroto, L, 1981). Berikut merupakan Tabel imbuhan awalan didalam bahasa Melayu Manado, dapat dilihat pada Tabel 2.2.

Tabel 2. 2 Prefiks (Awalan)

No	Prefiks	Contoh	Keterangan
State Islamic University of Sultan Syarif Kasim Riau	Ba-	Ba + Ombong = Baombong	- Membentuk Kata Sifat dari kata benda.
		Ba+ Photo = Bapoto	- Membentuk kata kerja dari kata benda.
		Ba + Pelang = Bapelang	- Membentuk Kata kerja dari kata sifat.
		Ba + Kuat = Bakuat	- Membentuk Adverbial dari kata sifat.
		Ba + tandu = Batandu	- Dipakai dalam kata benda dan mengandung makna mempunyai.
		Ba+karja = Bakarja	- Dipakai dalam kata kata kerja dan mengandung makna melakukan apa yang disebut dalam kata dasar.
		Bar + anak = Baranak	- Membentuk kata sifat dari kata benda.
State Islamic University of Sultan Syarif Kasim Riau	Bar-	Bar + bage = Barbage	- Membentuk adverbial dari kata kerja.
		Bar + dosa =	- Dipakai didalam kata benda dan mengandung makna mempunyai.

No	Prefiks	Contoh	Keterangan
1		Bardosa	
		Bar + guna = Barguna	- Dipakai didalam kata kerja dan mengandung dalam keadaan seperti yang disebut oleh kata kerja tersebut.
		Bar + dua = Bardua	- Dipakai didalam kata bilangan dan mengandung makna menjadi.
		Baku + siku = Bakusiku	- Membentuk kata resiprok (saling) dari kata benda.
2	Baku-	Baku + tinggi = Bakutinggi	- Membentuk kata kerja dari kata sifat.
		Baku + tunggu = Bakutunggu	- Membentuk makna resiprok (saling) dari kata kerja
		Baku+rebe = Bakurebe	- Membentuk makna resiprok (saling) dari kata keterangan.
		Ta + knop = Taknop	- Membentuk kata kerja dari kata benda.
3	Ta-	Ta + pinggir = Tapinggir	- Membentuk kata kerja dari kata sifat.
		Ta + ambor = Taambor	- Membentuk kata kerja dari kata adverbial.
		Ta + tidor = Tatidor	- Bermakna ketidak sengajaan dari kata kerja.
		Ta + butul = Tabutul	- Bermakna ketidak sengajaan dari kata sifat.
		Tar + maso = Tarmaso	- Berfungsi untuk membentuk makna ketidak sengajaan dari kata kerja.
4	Tar-	Tar+dorong = Tardorong	- Membentuk adverbial dari kata kerja.
		Tar+sangka = Tarsangka	- Membentuk kata benda dari kata kerja.
		Ka + Dua = Kadua	- Membentuk kata bilangan ordinal menjadi kata bilangan cardinal.
5	Ka-	Ka + Satu = Kasatu	Membentuk kata benda dari kata kerja.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

No	Prefiks	Contoh	Keterangan
1		Ka + dua = Kadua	
		Ka + Tiga = Katiga	
		Ka + Spuluh = Kaspuluh	
2	Pan-	Pan + toki = Pantoki	<ul style="list-style-type: none"> - Membentuk kata benda dari kata sifat. - Membentuk pelaku dari kata benda
		Pan+tutu = Pantutu	
		Pan + Cukur = pancukur	
		Pan + Dusta = Pandusta	
3	Pa-	Pa + Mabo = Pamabo	<ul style="list-style-type: none"> - Membentuk pelaku dari kata benda. - Membentuk kata benda dari kata sifat.
		Pa + marah = Pamarah	
		Pa + Tipu = Paniput	
4	Pam	Pam + Bae = Pambae	<ul style="list-style-type: none"> - Membentuk pelaku dari kata benda. - Membentuk kata benda dari kata sifat.
		Pam + Barani = Pambaranoi	
		Pam + Prop = Pamprop	
5	pang	Pang + kotor = pangkotor	Membentuk pelaku dari kata sifat
		Pang + sayang = pangsayang	

Sufiks

Imbuhan Akihran (Sufiks) digunakan dalam bahasa Melayu Manado terdiri dari (an- dan akang) (Ny.J.A.Karisoh Najoan, M.A Liwoso, Kinajati Djojuroto, L,

1981). Berikut merupakan tabel imbuhan akhiran didalam bahasa Melayu Manado, dapat dilihat pada Tabel 2.3.

Tabel 2. 3 Akhiran (Sufiks)

No	Suffiks	Contoh	Keterangan
1	-an	Tulis + an = tulisan	- Membentuk kata kerja dari kata benda
		Manis + an = manisan	- Membentuk kata benda dari kata sifat.
		Kilo + an = kiloan	- Membentuk adverbial dari kata benda
2	-akang	Wora + akang = woraakang	- Membentuk kata kerja dari kata sifat.
		Prop + a kang = propakang	- Membentuk kata kerja dari kata benda.
		Potong + akang= potongakang	- Mengandung makna tolong atau bantu

Reduplikasi (Perulangan)

Reduplikasi merupakan sebuah perulangan kata yang terdapat didalam sebuah kata (Ny.J.A.Karisoh Najooan, M.A Liwoso, Kinajati Djojurosoto, L, 1981). Berikut merupakan Tabel Reduplikasi didalam bahasa Melayu Manado, dapat dilihat pada Tabel 2.4.

Tabel 2. 4 Reduplikasi (Perulangan)

Reduplikasi	Contoh	Keterangan
Reduplikasi kata	Makang-makang	- Membentuk kata kerja dari kata sifat
	Rumah-rumah	- menyatakan makna jamak
	Bajalang-jalang	- Menyatakan pekerjaan yang berulang

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Kombinasi afiks

Kombinasi afiks didalam bahasa Melayu Manado terdiri dari (Ba+baku, Ta+baku dan Ta+ka)(Ny.J.A.Karisoh Najooan, M.A Liwoso, Kinajati Djojuroto, L, 1981). Berikut merupakan Tabel Kombinasi Afiks didalam bahasa Melayu Manado, dapat dilihat pada Tabel 2.5.

Tabel 2. 5 Kombinasi Afiks

No	Kombinasi Afiks	Contoh	Keterangan
1	Ba+ Baku	Ba+baku+kore = babakukore	Untuk menyatakan makna saling
2	Ta+ Baku	Ta+baku+sontong = tabakusontong	Untuk menyatakan makna saling dengan tidak sengaja.
3	Ta+ Ka	Ta+ka+huk = takahuk	Untuk menyatakan pasif yang tidak sengaja.
4	Pam + Ba	Pang+ba+Tare = Pambatare	Untuk menyatakan makna gemar atau suka
5	Pang+ Ba	Pang+ba+diang =	Untuk menyatakan makna gemar atau suka

Konfiks (Gabungan awalan dan akhiran)

Konfiks didalam bahasa Melayu Manado terdiri dari (ka-+-an dan Pan-+-an)(Ny.J.A.Karisoh Najooan, M.A Liwoso, Kinajati Djojuroto, L, 1981). Berikut merupakan Tabel Konfiks didalam bahasa Melayu Manado, dapat dilihat pada Tabel 2.6.

Tabel 2. 6 Konfiks

No	Konfiks	Contoh	Keterangan
1	Ka-+-an	Ka + basar + an = kabasaran Ka+ malas + an = kamalasan	- Membentuk kata benda dari kata sifat. - Membentuk kata benda dari kata dari kata kerja. - Membentuk kata benda dari

Hak Cipta Dilindungi Undang-Undang

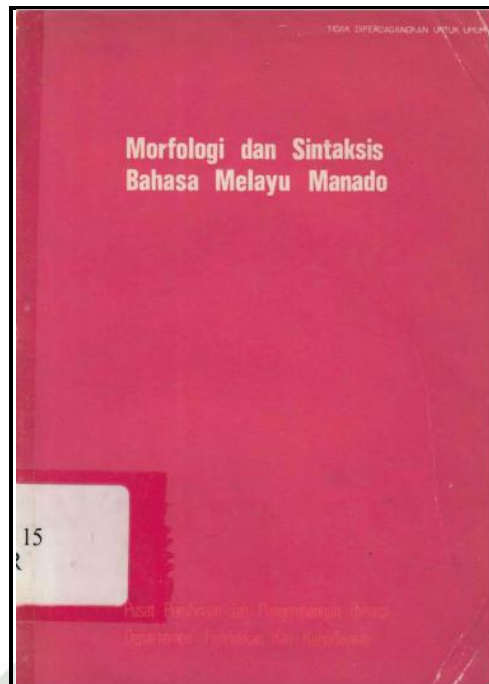
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

			kata keterangan. - Membentuk kata benda dari kata kerja.
	Pan- + -an	Pan + jual + an = penjualan Pan+ cari+ an = pancarian	- Membentuk kata benda dari kata kerja - Membentuk makna sesuatu dari kata dasar. -
	Pa- + -an	Pa+ lelang + an = palelangan Pa+Potong+ an=pamotongan	- Membentuk kata benda dari kata kerja - Membentuk makna sesuatu dari kata dasar
4	Pam- + -an	Pam + bunuh + an = pambunuhan Pam + baca + an = pambacaan	- Membentuk kata benda dari kata kerja - Membentuk makna sesuatu dari kata dasar
5	Pang- + -an	Pang + ajar + an = pangajaran Pang + Gali + an = penggalian	- Membentuk kata benda dari kata kerja - Membentuk makna sesuatu dari kata dasar

Penjabaran aturan tata bahasa atau morfologi Melayu Manado di atas didasarkan pada sumber buku “Morfologi dan Sintaxis Bahasa Melayu Manado”. Buku sumber morfologi dapat dilihat pada gambar 2.2.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 2.2 Sumber Morfologi

2.8 Perhitungan Akurasi (*Suggestion Adequacy*)

Perhitungan akurasi didasarkan pada pemeriksaan rekomendasi yang relevan dan akurat kepada pengguna untuk seluruh kata yang mengalami kesalahan ejaan. *Suggestion Adequacy* pemeriksaan ejaan seharusnya didasarkan pada *true negative* dan tidak pada *all negative*, karena tujuannya untuk menentukan seberapa baik pemeriksaan ejaan bisa menyarankan koreksi untuk kata-kata yang salah (Van Hyyssteen, 2004).

Pada penghitungan akurasi penelitian ini berlaku aturan sebagai berikut:

- Rekomendasi yang sesuai ada pada urutan pertama maka bernilai 1.
- Rekomendasi yang sesuai ada pada urutan 2 dan seterusnya maka bernilai 0.5.
- Tidak ada rekomendasi yang sesuai maka bernilai -0.5
- Tidak ada saran maka bernilai 0.

Akurasi dapat dihitung menggunakan rumus akurasi sebagai berikut :

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$SA = \sum_{k=0}^n \frac{k}{N_{Tn}}$$

2.9 Penelitian Terkait

Berikut beberapa penelitian terdahulu terkait normalisasi teks, dapat dilihat pada Tabel 2.7.

Tabel 2. 7 Penelitian Terkait

No	Peneliti	Judul	Metode	Hasil
1	Pratama & Pamungkas, 2016	Analisis Kinerja Algoritma Levenshtein Distance dalam Mendeteksi Kemiripan Dokumen Teks	<i>Levenshtein Distance</i>	Hasil simulasi menunjukkan bahwa penggunaan sorting sangat berpengaruh bagi algoritma Levenshtein distance. Hasil terbaik pada data set 1 ditunjukan pada proses yang menggunakan stopword removal, stemming, dan sorting sekaligus.
2	Nugraha & Rizqullah, 2019	Normalisasi Kata Tidak Baku yang Tidak Disingkat dengan Jarak Perubahan	<i>Levenshtein Distance & Longest Common Subsequence</i>	Hasil pengujian menunjukkan bahwa normalisasi dengan jarak Levenshtein mengungguli normalisasi dengan jarak Longest Common Subsequence (LCS) dengan selisih akurasi sebesar 8,34%.
3	Stezar Priyansyah & Renny Pradina	Normalisasi Teks Media Sosial Menggunakan Word2vec, Levenshtein Distance, dan Jaro-Winkler	<i>Dictionary Lookup dan Damerau-Levenshtein Distance</i>	Penelitian mendapatkan akurasi 25% pelatihan model word2vec dan pengujian sampel terbaik pengujian data mendapatkan akurasi

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

No	Peneliti	Judul	Metode	Hasil
		Distance		terbaik adalah 75.9% dengan threshold sebesar 70%.
	Prananda Antinasari, Rizal Setya Perdana, M. Ali Fauzi, 2017	Analisis Sentimen Tentang Opini Film Pada Dokumen <i>Twitter</i> Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku	<i>Levenshtein Distance & Naïve Bayes</i>	Berdasarkan hasil pengujian yang telah dilakukan didapatkan akurasi tertinggi dengan nilai <i>accuracy, precision, recall</i> , dan <i>f-measure</i> sebesar 98.33%, 96.77%, 100%, dan 98.36%.
	Ritika Mishra, Navjot Kaur	A Survey of Spelling Error Detection and Correction Techniques	<i>Spell Checker, Error Detection, Error Corection</i>	Hasil dari penelitian ini adalah Makalah ini menyajikan studi tentang berbagai Ejaan Teknik koreksi. Jadi, di masa depan, kami akan merancang dan mengimplementasikan Pemeriksa Ejaan baru untuk Bahasa India.
6	Mohammad Ali Elmi dan Martha Evens	Spelling Correction Using Context*	<i>Spelling Correctin</i>	Hasil dari penelitian ini adalah Algoritma koreksi ejaan memperluas tiga cara mencocokkan algoritma dan penawaran dengan masalah kata yang disingkatan. Masalah itu bisa diatas dengan gabungan antara parsing dan spelling correction. Tujuan pertama dari penelitian ini adalah untuk mendeteksi kesalahan yang terjadi ketika kata-kata berubah menjadi lain kata-kata dalam

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

No	Peneliti	Judul	Metode	Hasil
				leksikon.
	Hany Hasan dan Arul Menezes(Hassan & Menezes, 2013)	Social Text Normalization using Contextual Graph Random Walks	Contextual Graph Random Walks	Hasil dari penelitian ini adalah dengan menggunakan sistem sebagai preprocess-langkah untuk sistem terjemahan mesin yang meningkatkan kualitas terjemahan sebesar 6%. Sebagai perpanjangan dari pekerjaan ini, juga akan memperluas pendekatan untuk menangani ke lebih banyak Bahasa.
8	Dewi Rosmala, Zulfikar Muhammad Risyad (2017)	Algoritma <i>Levenshtein Distance</i> dalam aplikasi pencarian kata isu di Kota Bandung pada Twitter.	<i>Levenshtein Distance</i>	Hasil dari implementasi tersebut dapat disimpulkan bahwa penggunaan Algoritma <i>Levenshtein Distance</i> dalam Aplikasi Pencarian Kata Isu di Kota Bandung Pada Twitter yang mengandung kata dengan kesalahan ejaan menjadi kata kunci yang kemudian dimasukkan ke dalam daftar kategori isu yang terdapat di Pemerintah Kota Bandung dengan tingkat akurasi 100%
	Richard Khou(Khoury, 2015)	<i>Microtext Normalization Using Probaby-Phonetically-Similar Word Discovery</i>	<i>Kombinasi Algoritma Levenshtein Didtance, Phonetic Tree dan Dictionary-Based</i>	Pengujian menggunakan 2608 OOV Word dengan akurasi normalisasi sebesar 80%.
10	Benisius 2017	Sistem	<i>Levenshtein</i>	Hasil simulasi



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

No	Peneliti	Judul	Metode	Hasil
		pengkoreksian kata kunci dengan menggunakan metode <i>Levenshtein Distance</i>	<i>Distance</i>	menunjukkan Metode Levenshtein Distance dapat diterapkan untuk melakukan proses pengkoreksian kata kunci pada aplikasi pencarian berbasis web. Sistem dapat memproses inputan kata kunci dalam jumlah yang banyak dengan durasi waktu eksekusi yang tergantung dari kompleksitas kata kunci yang diinputkan.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB III

METODOLOGI PENELITIAN

Metodologi penelitian merupakan langkah-langkah yang ditempuh untuk melakukan penelitian dan mendapatkan hasil dari rumusan masalah. Langkah-langkah penelitian dapat dilihat pada Gambar 3.1 sebagai berikut:



Gambar 3.1 Flowchart Metodologi Penelitian

3.1 Perumusan Masalah

Perumusan masalah adalah proses penentuan akar masalah yang akan diselesaikan dalam penelitian ini. Tujuan adanya perumusan adalah agar penelitian tidak keluar dari akar permasalahan. Dengan adanya akar permasalahan yang telah dicantumkan pada latar belakang, maka perumusan masalah dalam

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

penelitian ini adalah “Bagaimana menerapkan algoritma *Levenshtein Distance* untuk normalisasi kata pada bahasa Melayu Manado”.

3.2 Studi Literatur

Studi literatur adalah pengumpulan informasi yang dapat mendukung teori dalam penelitian yang dilakukan. Sumber teori pendukung berupa buku dan jurnal. Teori yang dibutuhkan adalah teori seputar normalisasi kata, *Levenshtein Distance* serta teori mengenai *text preprocessing*.

3.3 Pengumpulan Data

Pengumpulan data merupakan tahapan selanjutnya yang akan dilakukan oleh penulis untuk mendapatkan data-data yang dibutuhkan. Data yang dikumpulkan untuk melanjutkan dan mendukung penelitian ini diperoleh dari:

1. Kamus bahasa Melayu Manado yang dijadikan sumber kata dasar bahasa Melayu Manado. Kata yang terdapat dalam kamus ini akan dimasukkan kedalam basis data (Rahmad, 2019).
2. Tata bahasa bahasa Melayu Manado digunakan untuk mencari kata berimbuhan. Morfologi didapat dari penelitian sebelumnya mengenai *stemming* bahasa Melayu Manado (Rahmad, 2019).
3. Data uji berbahasa Melayu Manado diperoleh dari media sosial *twitter* dengan nama akun @TheBAKUSEDU.

3.4 Analisa Algoritma

Pada tahap ini penulis akan melakukan analisa terhadap data-data yang telah diperoleh. Analisa dilakukan untuk mengetahui alur proses normalisasi. Analisa yang dilakukan untuk menerapkan metode adalah:

3.4.1 Analisa Kebutuhan Data

Pada tahapan ini akan dilakukan analisa terhadap kebutuhan data untuk melaksanakan penelitian. Data yang dibutuhkan adalah data kata dasar dan data berupa *tweet* berbahasa Melayu Manado sejumlah 100 *tweet* yang

berkemungkinan terdapat kesalahan penulisan dari sebuah akun bernama “TheBAKUSEDU”.

3.4.2 Analisa Tahapan

Tahap pertama pada proses normalisasi, yakni melakukan *cleaning*. Pada tahap ini akan dilakukan proses penghapusan karakter yang bukan merupakan huruf. Contohnya adalah menggunakan karakter koma (,) atau penambahan-penambahan *emoticon* di dalam sebuah teks. Tahap berikutnya merupakan *case folding* terhadap teks yang didapat dari media sosial, sehingga seluruh teks yang semula berkemungkinan memiliki huruf kapital akan dirubah menjadi huruf kecil. Tahap berikutnya merupakan *tokenizing*, pada tahap ini teks tersebut akan dikonversi menjadi *string*, dengan kata lain kalimat dalam teks akan dirubah menjadi kata per kata. *String* yang dihasilkan dari proses sebelumnya akan mengalami proses *stemming* yaitu proses penghilangan imbuhan, sehingga *string* tersebut akan berbentuk kata dasar. Kata dasar ini akan beracuan pada kamus bahasa Melayu Manado. Jika kata yang dicari tidak ditemukan pada kamus, maka akan dilakukan normalisasi. Pada penelitian ini, normalisasi akan menggunakan metode *Levenshtein Distance*.

Levenshtein Distance adalah teknik menghitung jarak kemiripan dua buah *string*. Tahapan yang dilakukan dalam *Levenshtein Distance* berupa penambahan, penghapusan dan penggantian karakter dalam *string*. Setiap kata yang tidak memiliki kecocokan dalam kamus stemming akan melalui proses normalisasi dengan menghitung jarak kemiripan kata tersebut dengan kata yang ada dalam kamus. *Output* yang dihasilkan adalah rekomendasi kata yang memiliki jarak kemiripan terkecil.

3.5 Perancangan

Pada tahapan perancangan ini, akan dibuat aplikasi normalisasi bahasa Melayu Manado berdasarkan hasil analisa dari langkah-langkah sebelumnya. Perancangan yang akan dilakukan berupa perancangan basis data, perancangan antarmuka dan *flowchart*.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.5.1 Perancangan Basis Data

Perancangan basis data adalah proses membangun sebuah penyimpanan data yang dibutuhkan untuk proses normalisasi bahasa Melayu Manado. Perancangan ini dilakukan untuk mengumpulkan informasi-informasi yang disimpan secara sistematis di dalam komputer.

3.5.2 Perancangan Antarmuka

Perancangan antarmuka adalah pembuatan tampilan aplikasi yang akan dibangun. Perancangan antarmuka bertujuan agar pengguna aplikasi dapat mengerti dalam menjalankan aplikasi normalisasi bahasa Melayu Manado.

3.5.3 Flowchart

Flowchart atau diagram alir adalah gambaran langkah-langkah yang dilakukan dalam melakukan penelitian. Flowchart digambarkan dengan simbol arahan sesuai dengan jalan analisa.

3.6 Implementasi

Pada tahapan implementasi dan pengujian merupakan tahapan yang dilakukan setelah tahapan analisa dan perancangan selesai dilakukan. Pada tahap ini menjelaskan secara rinci bagaimana penerapan sistem yang telah selesai analisa dan dirancang sebelumnya. Adapun kebutuhan alat dan pendukung untuk proses implementasi yaitu perangkat keras (*hardware*) dan perangkat lunak (*software*) sebagai berikut.

Perangkat keras yang dibutuhkan, yaitu :

Processor : AMD 4 Core A10-7400P, up to 3.4GHz
Memory : 4.00 GB

Perangkat lunak yang dibutuhkan, yaitu :

Platform : Windows 10
 Bahasa Pemrograman : PHP
DBMS : MySQL
Server : Apache2



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

3.7 Pengujian

Pengujian (*testing*) yaitu uji coba apakah tingkat akurasi sesuai dengan yang sebenarnya atau tidak. Setelah melakukan implementasi maka harus dilakukan pengujian apakah aplikasi tersebut telah sesuai dengan yang diharapkan. Penelitian ini diuji dengan *Suggestion Adequacy* dan pengujian oleh validator. Validator adalah orang yang memvalidasi data uji dan hasil pengujian. Validator pada penelitian ini adalah seorang dosen UNSRAT Manado, yaitu Bapak Fredy M.S.B Wowor orang yang berlatar belakang Manado dan dapat menggunakan bahasa Melayu Manado.

3.8 Kesimpulan dan Saran

Tahapan kesimpulan dan saran adalah tahapan terakhir dari penelitian. Pada tahap ini berisi kesimpulan yang dapat ditarik berdasarkan hasil dari penelitian. Sedangkan saran memberikan masukan untuk mengembangkan penelitian selanjutnya.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB VI

Kesimpulan dan Saran

6.1 Kesimpulan

Berdasarkan penelitian yang dilakukan didapatkan kesimpulan bahwa metode *Levenshtein Distance* telah berhasil diterapkan dengan akurasi yang dihasilkan dari penerapan metode *Levenshtein Distance* untuk normalisasi kata typo Melayu Manado bernilai 35.8%.

6.2 Saran

Dari hasil penelitian yang dilakukan, dapat dilakukan pengembangan ataupun dilakukan penelitian lanjutan dengan beberapa saran. Saran yang dapat diberikan adalah Melakukan normalisasi kalimat bahasa Melayu Manado dengan metode lanjutan N-Gram Language Model sehingga kata yang mengalami kesalahan pengetikan (*typo*) dapat di normalisasikan sesuai dengan konteks kalimat.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR PUSTAKA

- Amamanda, R., Suhery, C., & Brianorman, Y. (2016). Aplikasi Pendeteksi Plagiat Terhadap Karya Tulis Berbasis Web Menggunakan Natural Language Processing dan Algoritma Knuth-Morris-Pratt. *Jurnal Coding, Sistem Komputer Untan*, 04(1), 33–44.
- Asri, H. (2014). Verba dalam Bahasa Melayu Manado. *Sawerigading*, 20(2), 311–320.
- Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, 9(2), 48–57. <https://doi.org/10.1109/MCI.2014.2307227>
- Darujati, C., & Bimo Gumelar, A. (2016). Pemanfaatan Teknik Supervised Untuk Klasifikasi Teks Bahasa. *Jurnal Link*, 16(February 2012), 8.
- Devianty, R. (2017). Bahasa Sebagai Cermin Kebudayaan. *Jurnal Tarbiyah*, 24(2), 226–245.
- Hassan, H., & Menezes, A. (2013). Social text normalization using contextual graph Random Walks. *ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, 1, 1577–1586.
- Khoury, R. (2015). *Microtext Normalization using Probably- Phonetically-Similar Word Discovery*. 384–391.
- Lyras, D. P., Sgarbas, K. N., & Fakotakis, N. D. (2007). Using the levenshtein edit distance for automatic lemmatization: A case study for modern greek and english. *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 2, 428–435. <https://doi.org/10.1109/ICTAI.2007.41>
- Mujilahwati, S. (2016). Pre-Processing Text Mining Pada Data Twitter. *Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016*(Sentika), 2089–9815.
- Najjoan, K., Liwoso, Djojuroto, K., & Kembuan, L. (1981). *Morfologi dan Sintaksis Bahasa Melayu Manado*.
- Nugraha, I. G. B., & Rizqullah, R. D. (2019). Normalisasi Kata Tidak Baku yang Tidak Disingkat dengan Jarak Perubahan. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)*, 8(3), 218.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

<https://doi.org/10.22146/jnteti.v8i3.516>

Pratama, B. P., & Pamungkas, S. A. (2016). Analisis Kinerja Algoritma Levenshtein Distance dalam Mendeteksi Kemiripan Dokumen Teks. *Jurnal "LOG!K@,"* 6(2), 131–143.

Pransya, S. (2017). *Normalisasi Teks Media Sosial Menggunakan Word2Vec, Levenshtein Distance, dan Jaro-Winkler Distance.*

Purwadi. (2011). *Purwadi.*

Putri, E. (2014). Penerapan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes. *Penerapan Text Mining Pada Sistem Klasifikasi Email Spam Menggunakan Naive Bayes,* 2(3), 73–83.
<https://doi.org/10.12928/jstie.v2i3.2877>

Rahate, P. M., & Chandak, M. B. (2019). Text normalization and its role in speech synthesis. *International Journal of Engineering and Advanced Technology,* 8(5 Special Issue 3), 115–122.
<https://doi.org/10.35940/ijeat.E1029.0785S319>

Sharma, D. (2012). Nuove tecniche di istoterapia. *International Journal of Applied Information Systems (IJ AIS),* 4.

Subagia, R., Alit, R., & Akbar, F. A. (2020). Pengujian white box pada sistem informasi monitoring skripsi program studi informatika. *Jurnal Informatika Dan Sistem Informasi (JIFoSI),* 01(2), 539–547.

Widianto, E. (2018). Pemertahanan Bahasa Daerah melalui Pembelajaran dan Kegiatan di Sekolah. *Jurnal Kredo, (1)* 2, 1–13.

Zulkifli, U. C., & Suadaa, L. H. (2019). *Pengembangan Modul PreprocessingTeks untuk Kasus Formalisasi dan Pengecekan Ejaan Bahasa Indonesia pada Aplikasi Web Mining Simple Solution (WMSS).* (2), 95–103.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Lampiran A

Data Validator

Nama	: Fredy M S B Wowor
Tempat/ Tgl Lahir	: Tomohon, 04-03-1977
Jenis Kelamin	: Laki-Laki
Alamat	: Dusun 1 Talikuran No. 35 Kecamatan Sonder, Sulawesi Utara
Agama	: Kristen
Status Perkawinan	: Belum Kawin
Pekerjaan	: Pegawai Minat Bakat Kantor Pusat UNSRAT Manado Dosen Luar Biasa Di Jurusan Sastra Indonesia Fakultas Ilmu Budaya UNSRAT Manado



Lampiran B

Pengujian Data Uji

No	Data Uji	Kata Typo	Rekomendasi Sistem	Hasil Validator	Nilai
11.	Orng bakudpa di sana	orng	orang	orang	1
		bakudpa	bakudapa ,bakudepa,bakuopa,bakup a,bakupa	bakudapa	1
12.	lebe bae putus deng senyuman, dari pada mo batahang kong malele aer mata trus	-	-	-	
13.	yang so harum jam bgini mar depe laste cuma badiam di leput	bgini	bagini,bini	bagini	1
14.	yang ba jam bgni cuma	begini	bagini	bagini	1
15.	cwek sapa suka jadi deng qta	cwek	bek, blek, ces, cet, ceweq , zwak	ceweq	0.5
		qta	ada, atas, ato, bata, bia, kita , weta	kita	0.5
16.	ngana dang slh stunya bakusedu yang ba jam bgni Cuma	slh	ali, alo, bla, bla, bli, boh, ela, sala	sala	0.5
17.	qta biar pasaran mar nda mura g da bobral hey cwek sapa suka jadi deng qta	murah	marah, mura, mura , muran	mura	0.5
		bobral	baobral , obral	baobral	1
18.	menurut cweq yg so sering patah hati cowo di dunia ini cuma ada jenis banci	menurut	baturut , benaut, katurut, manurip, mendur, turut	baturut	1
		yg	af,as,bu,ca,ci,co, yang , yong	yang	0.5
		patah	pata , patai, patau, patoh, pazah	pata	1
19.	di dunia qta p mama dia da fase lahir	p	pe , pi, pu	pe	1
20.	bus ngana pe kamar pe skali qta so begitu noh qt p kisah cinta	qta	ada, atas, ato, bata, bia, kita , weta	kita	0.5
		qt	af, as,ato, bu, ca, cet, x	kita	-0.5

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1.	cwe harap waspada skarang so jam cwo brengsek baambor potas	cwe	ewe	ceweq	-0.5
2.	hey cantik so boleh tidor jo yang lupa eso pagi kse tau pa at tu mimpi dri mo syair	hey	hei ,her	hei	1
		kse	kase , kuse, se	kase	1
		qt	af, as, ato, bu, ca, cet, x	kita	-0.5
3.	selamat pagi mencari	selamat	slamat	slamat	1
		mencari	pancari, pencar	mancari	-0.5
24.	yang mo pigi skg mo dapa keuntungan dapa kerja dapa pasangan	skg	eke, eko, ik, ika, iko, oki, suku	skarang	-0.5
		keuntungan	kauntungan	kauntungan	1
25.	tu yg so klar lebe bae pi coba cari kerja jang cuma mantn p nda m dapa ngni paling saki hati	mantn	maitan, manta, mantaan, matan	mantan	-0.5
		p	pe , pi, pu	pe	1
		nda	ada,nua,nya	nyandaq	-0.5
		ngni	ngoni	ngoni	1
26.	jam bagini kong mantan selamat pagi langsung lome ng kwa ni badan	mantn	maitan, manta, mantaan, matan	mantan	-0.5
		ulg	ali, alo,bla, bli, bulu, vla	ulang	-0.5
27.	panas betul	-	-	-	
28.	panas bagini paling sadap pansi d kulkas	d	di , do	di	1
29.	brapa banyak yg ada di skarang	yg	af, as, bu, ca, ci, co, yang , yong	yang	0.5
30.	tahan trus jo mar klu so nd tahan lebe bae putus	tahan	taan,taban,tadan,t ahak, tahang ,taha p	tahang	0.5
		klu	kalu	kalu	1
		nd	ne,ni,no,nod	nyandaq	
31.	ong bilng qt nimbole jadi contoh for org laeng	orng	orang	orang	1
		bilng	biang, bilang	bilang	0.5
		org	oro	orang	-0.5

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

© Hak Cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

32.	kalu samua laki itu sama kyapa ni cwek lama skali for mo bapilih	kyapa	kiapa	kiapa	1
		cwek	bek, blek, ces, cet, ceweq , zwak	ceweq	0.5
		bapilih	bapili	bapili	1
33.	hey ini so mulai ujang jang bermain ujang	hey	hei her	hei	1
		mulai	molai	molai	1
		bermain	barkain, barmai, barmaian, barmaing , barmaid, beraan, birman, birman, jerman	barmaing	0.5
34.	zaman skarang msh taon so hamil	zaman	faman, jaman , ram an, zahan, zalan	jaman	0.5
		msh	as, asi, asu, boh, eso, isi, masi	masi	0.5
35.	saki hati itu cuma sementara	sementar a	samantara	samantara	1
36.	paling gmpang mo bilang mar susah mo beking	gmpng	gampang	Ada	1
37.	heyw cwee pi batona jo dang deng	heyw	eya, hede, hei , hele, her, heri	hei	0.5
		batona	banona, batana, bat oa, batoha, batoka, batola, batonaq , batono, batuna	batonaq	0.5
38.	tanya pa sebe deng ajus kalu apa tu dorang p kesalahan terbesar p drg p hidop dorang cuma manangis kong bpolo pa ta apa maksud	btanya	banya, banya, bata nya , tanya, tanya	batanya	0.5
		kesalaha n	kasalaan	kasalaan	1
		bpolo	bapolo polo	bapolo	1
39.	jaman skarang p batrei tu lebe tahan	tahan	taan, taban, tadan, tahak, tahang , tahap	tahang	0.5
40.	asal yang qt da beking kong so ta lebe sayang p ngna kangen p ngana makang	da	ada , ca, dak, dan, di, do, ja, ra	ada	1
		ta	ca, ja, ra, tai, tau, tax	kita	-0.5



41.	am bagni sdap skli m liat bakusedu dpe isi duri samua	bagni	bagai, bagini ,bago ni,bagui,bani	bagini	0.5
		sdap	sadap ,slap,snap, stap	sadap	1
		m	mo , om	mo	1
		dpe	depe , pe,upe	depe	1
		skli	sili, skali	skali	0.5
42.	percuma ja batona kalu tu peda tetap bkarat	batona	banona,batana,bat oa,batoha,batoka, batola, batonaq , batono,batuna	batonaq	0.5
		bkarat	bakarat , karat	bakarat	1
43.	oh jadi ngn pe paytua so tidor dank ini ada bukti	ngn	nin	ngana	
		paytua	paitua ,pamtua,pa ntua,patua	paitua	1
		dank	dak,dan, dana, dang	dang	0.5
44.	tiap pagi qta qt musti bangun pagi pi balia cwe gaga tanpa tidor syg di luar msh dingin tidor ulang	qta	ada, atas, ato, bata, bia, kita , weta	kita	0.5
		qt	af, as,ato, bu, ca, cet, x	kita	0.5
45.	cwe itu kalu bablg cari akg paitua dang pa qta itu ada barangkali dia suka p ng dia suka pa ng p tmng	cwe	ewe	ceweq	-0.5
		bablg	babla,babli	babelang	-0.5
		ng	ne,ni,no	ngana	-0.5
		akg	abu,ada,ade,ado,a er,af,aju, akang	akang	0.5
		p	pa, pe ,pi,pu	pe	0.5
		tmng	dang,deng,dong,h ong, tamang	tamang	0.5
46.	kalu hari ini ujang kras brarti au doa so jadi	-	-	-	
47.	ujang jo kras sampe esok	esok	enok, eso ,wesok	eso	0.5
48.	malam minggu bagini kong ng p paitua nda dtg di rumah kuma lantaran rinte lebe bae	malam	malap,malas	malang	-0.5
		ng	ne,ni,no	ngana	-0.5
		p	pa, pe ,pi,pu	pe	0.5

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

49.	kalu paitua nd datang cuma antaran rinte berarti ngni itu cuma depe maitua kalu musim panas klu musim ujang brtti dia ada laeng	nda	ada,nua,nya	nyanda	-0.5
		dtg	ato,dak,dan,dang, datu, utu	datang	-0.5
		nd	ne,ni,nonod	nyanda	-0.5
		ngni	ngoni	ngoni	1
		klu	kalu	kalu	1
50.	daripada sakihati ba ha org batona lebe bae pi bermain sabong jo ngoni di kamar mandi	brtti	arti, baarti ,barai	baarti	0.5
		batona	banona,batana,bat oa,batoha,batoka, batola, batonaq , batono,batuna	batonaq	0.5
		bermain	Barkain,barmai,ba rmaian, barmaing	barmaing	0.5
51.	biar jo qt jomblo daripada sllu bdusu for mo bakubale	qt	af, as,ato, bu, ca, cet, x	kita	-0.5
		sllu	balu,blou,bolu,bul u,dolu, salalu ,tolu	salalu	0.5
		bdusu	badusu , busu, dusu	badusu	1
52.	kalu dulu org pi maso grja for m dengar firman skg pi maso grja spy boleh m update status at church ato sunday service kacauu	org	oro	orang	-0.5
		grja	gereja	gereja	1
		m	mo ,om	mo	1
53.	bangun tidor gt bakaca gt masih ganteng gt tidor ulang	gt	go	gaga	-0.5
54.	sapa mo btanya silahkan	btanya	banya,banya, bata nya ,tanya,tanya	batanya	0.5
55.	somo k deng ng bakusedu so brpa tengkorak	k	ik	kah	-0.5
		ng	ne,ni,no	ngana	-0.5
		brpa	bapa,barpa, brapa	brapa	0.5
56.	bakusedu mo tanya min pe umur skrg brpa	skrg	ser,sero,siru,skel,s uru	skarang	-0.5
		brpa	bapa,barpa, brapa	brapa	0.5
57.	bakusedu so ada maitua ngna	ngna	nana, ngana ,nina, nona	ngana	0.5



58.	kweng bakusedu apa yg min lakukan kalau cinta pertama datang ulang	kweng	kaeng, kaweng	kaweng	0.5
		dtang	dang, datang , diang, itang, utang	datang	0.5
59.	it p sebe dg ajus p anak bakusedu sapa ngana	p	pa, pe ,pi,pu	pe	0.5
		dg	di,do	deng	-0.5
60.	bakusedu ng sekolah dmna	ng	ne,ni,no	ngana	-0.5
		dmna	dana	dimana	-0.5
61.	kalo da orang se php in nga bgemna	bgemna	baeman,baemanan ,bamana,barmana	bgemana	-0.5
		kalo	alo,falo,kaalo,kab o,kaco, kalu	kalu	0.5
62.	bukang qt yg bajingan mar qt p bu rung	qt	af, as,ato, bu, ca, cet, x	kita	-0.5
		p	pa, pe ,pi,pu	pe	0.5
63.	bakusedu so brp perawan min da se ambor lohang	brp	bae,bak,bal,ban,ba o, brapa	brpa	0.5
		lohang	lobang ,lowang	lobang	1
64.	bakusedu kalo sekarang matahari ada brapa	sekarang	skarang	skarang	1
65.	sebe bakusedu min sapa tu	-	-	-	
66.	bakusedu ngana pe gigi da brapa	-	-	-	
67.	bantal polo bakusedu da deng sapa skrg	skrg	ser,sero,siru,skel,s uru	skarang	-0.5
68.	rumah dudu sudah bakusedu di mana ba apa so makang	-	-	-	
69.	bakusedu suka bkng apa klu tengah mlm	bkng	akang,bakang bakong,ban, bekeng	bekeng	0.5
		klu	kalu	kalu	1
		mlm	ali, alo,bla,bla,fam	malang	-0.5
70.	dh bakusedu so mkang ngna	ngna	nana, ngana ,nina, nona	ngana	0.5
		mkang	akang,ikang,kang, makang	makang	0.5
71.	anya skali bakusedu ngana p maytua ada brp	maytua	maitua	maitua	1

Hak Cipta Ditanggung UIN Suska Riau

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

2.	main baju popi nga pe apa main atau main cwek	p	pa, pe ,pi,pu	pe	0.5
		nga	inga,noga,nua,nya	ngana	-0.5
		cwek	bek, blek,ces, cet, ceweq , zwak	Ceweq	0.5
		main	kain,mai,maian, m aing ,mait	maing	0.5
3.	lapan anam brenti jo bahugel lanjut jo bagate bakusedu banya skali	bahugel	bahagel	bahagel	1
74.	loreng putih ato itang ngana bakusedu	-	-	-	
75.	blh asal ng jadi dg qt	ng	ne,ni,no	ngana	-0.5
		blh	bla,bli,boh	boleh	-0.5
		qt	af, as,ato, bu, ca, cet, x	kita	-0.5
		dg	da,di,do	deng	-0.5
76.	nintau lubang apa paling kacili bakusedu	lubang	lobang	lobang	1
77.	bajalan kong baku pegang tangan itu wajar katu kalu batona mar nda perlu toh mo baku pegang tangan b jln sampe di wc	bajalan	babalan,bagalan,b ajaan,bajaan,bajae an, bajalang ,bazal an	bajalang	0.5
		bjln	baen,bal,balan,bal e,balu,bulu	bajalang	-0.5
78.	alasan qt binci k skola bangun pagi banya pr karna qt p skola nyanda hogwarts	qt	af, as,ato, bu, ca, cet, x	kita	-0.5
79.	ajus kyp sampe skarang mama taru kira ng rupa nda ada maitua qta mama qt mo cari blanga dulu baru mo cari depe tukang momasa	ng	ne,ni,no	ngana	-0.5
		qta	ada, atas, ato, bata, bia, kita , weta	kita	0.5
		kyp	kep,kop,kup	kiapa	-0.5
		momasa	bamasa,kamasa,k omasi,masa,molas , tomasan	masa	-0.5
80.	ata mo cemburu pa ng	ng	ne,ni,no	ngana	-0.5
81.	ang tako jatuh cinta ada yg lebe dapa tako jatuh cinta pa org yg salah	jatuh	batoh,batu,batua,b atuk,batus,datu,jat on, jatung	jatung	0.5
82.	at so bdusta kalu qt blg qta nda rindu p ngana	bdusta	badusta ,dusta	badusta	1
83.	amat pagi kalu ada mimpi	bbilang	babilang ,bailing,	babilang	1

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

	bagus bbilang		bilang		
84.	beking rupa ruangan kelas pasti cpt mo t tidor	ttidor	tatidor ,tidor	tatidor	1
85.	alasan cwe manangis senang sedih tanpa alasan	cwe	ewe	ceweq	-0.5
86.	masa ley kwa cwo mo bjalang dg cwe kong nda ada doi malo do	bjalang	bajalang , jalang	bajalang	1
87.	dululu org bilang cwe itu cuma butuh ng p waktu atensi dg kasih sayang mar skarang bole so ta tambah akg tu depe nama doi	org	oro	orang	-0.5
		cwe	ewe	ceweq	-0.5
88.	tidor jo tre samua cwe so ada paitua	cwe	ewe	ceweq	-0.5
89.	nda penting mo cari cwe kaya yang penting dia boleh beking tatawa p ngana	cwe	ewe	ceweq	-0.5
		p	pa,pe,pi,pu	pa	1
90.	pasangan yang tau cara mo beking torang bahagia dgn cara yg torang suka	torag	torang , torar	torang	1
91.	contoh bayar akang makang	contoh	conto ,lontoh,pont oh	conto	1
92.	laki pangmaraju itu pasti kecil suka main ceklen ato baju popi	main	kain,mai,maian, maing ,mait	maing	0.5
93.	jadi laki itu nimbole cengeng nda malo sese so bakumis	nda	ada,da,nua,nya	nyandaq	-0.5
94.	yang suka main main akang cewe pasti kecil suka main baju popi	main	kain,mai,maian, maing ,mait	maing	0.5
		cewe	cece,cele,cere, ceweq ,ewe	ceweq	0.5
95.	cewe yang suka main main akang pasti kecil perna anjing gigi	perna	Pera, pernah	pernah	0.5
96.	kap pi acara kaweng ni org na sllu ba tanya kapan	org	oro	orang	-0.5
		sllu	balu,blou,bolu,bulu,dolu, salalu	salalu	0.5
		qt	af, as,ato, bu, ca, cet, x	kita	-0.5
97.	cpt pulang jang ajus beking tanpa kos d p kamar saya diculik	cpt	cet,cit	cepat	-0.5
		kos	aos,kaos, kas,koa	kos	-0.5
98.	cabunuh kekerlak jgn buang depe bangkai kse biar jo dstu	kekerlak	kakerlak	kakerlak	1

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

99.	supaya jadi for depe tamang	bangkai	angka,baangka,bagai,bangisi, bangke	bangke	0.5
		kse	kase ,kuse,se	kase	1
		dstu	datu,dotu	disitu	-0.5
		jgn	ban,bon,dan,den,en,go	jangang	-0.5
100.	cwe skarang rupa so lebe aneh d m jadi gebetan mar susah skali apalagi jadi maitua	cwe	ewe	ceweq	-0.5
		skrng	pangkan,sangian,sangka,sangkaan,sangkoan	skarang	-0.5
		d	da ,di,do	da	1
		m	mo ,om	mo	1
101.	untung cinta itu buta coba kalu dia bole balia dorang yg sube kasiang mana mo laku	yg	af,as,bu,ca,ci,co, yang , yong	yang	0.5
102.	percuma bilang sayang pa qta mar ngana tetap pilih pa dia	percuma	barcuma, pacuma ,pamcuma,pancum a	pacuma	0.5
		qta	ada,atas,ato,bata,bia, kita ,weta	kita	0.5
103.	bilang sayang bilangleh cinta suruh datang tako setang	bilang	baanleh,baranleh, bilang ,bilangan	bilang	0.5
		suruh	suru	suru	1
104.	bilang sayang bilangleh cinta suruh datang tako setang	bilang	baanleh,baranleh, bilang ,bilangan	bilang	0.5
		suruh	suru	suru	1
105.	kalu ada pasangan jangan ika kuat pake bh saja nda nyaman kalu talalu kuat	bh	boh,bu	beha	-0.5
		nda	ada,da,nua,nya	nyandaq	-0.5
106.	kalu ngn baru jadian denk owo yg baru putus musti ika kuat pa dia kalu kase los pasti babale dia pa dp mantan	ngn	nin	ngana	1
		denk	dek,den, deng	deng	0.5
107.	so suka mo bobo mar nda ada orang yg suruh bobo nasib nasib	nda	ada,da,nua,nya	nyandaq	-0.5
		yg	af,as,bu,ca,ci,co, yang , yong	yang	0.5
		nasib	nasip	nasip	1
108.	ong saki sapa mo urus	-	-	-	



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

08.	qta pe dompet itu kurang sama deng bawang qta buka langsung ta manangis	qta	ada, atas, ato, bata, bia, kita , weta	kita	0.5
		ta	ca,da,ja,pa	kita	-0.5
09.	batona spya langgeng bacirita sama dg tamang bae bermain sama dg anak ado argumen sama dg org tua saling melindungi sama dg kaka ade	batona	banona,batana,bat oa,batoha,batoka, batola, batonaq , batono,batuna	batonaq	0.5
		spya	saya	supaya	-0.5
		dg	da,di,do	deng	-0.5
		ado	abu,ada,ade,ado,a due,gadu	adu	-0.5
110.	hal paling sedi lia ajus manangis	-	-	-	